ABSTRACT SYMPOSIUM NAME: DARPA Make-It Program: Automating Small Molecule Route Design, Optimization & Synthesis (Invited) (Oral)

ABSTRACT SYMPOSIUM PROGRAM AREA NAME: COMSCI

CONTROL ID: 2971032

PRESENTATION TYPE: Oral Preferred : Consider for Sci-Mix

TITLE: Automated System for Knowledge-based Continuous Organic Synthesis: Data-driven pathway design and validation

AUTHORS (FIRST NAME, LAST NAME): Connor W. Coley[1], Pieter Plehiers[1], Wengong Jin[2], Hanyu Gao[1], Yuran Wang[1], John Schreck[3], Kyle J. Bishop[3], Regina Barzilay[2], Tommi Jaakkola[2], William H. Green[1], Klavs F. Jensen[1]

INSTITUTIONS (ALL):

1. Chemical Engineering, MIT, Cambridge, MA, United States.
2. Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA, United States.
3. Chemical Engineering, Columbia University, New York, NY, United States.

ABSTRACT BODY:

Abstract: Advances in laboratory automation promise to decrease the manual effort of synthesis, but determining how to synthesize a compound currently requires time and effort investment from expert chemists. To achieve full autonomous chemical synthesis, one must have robust synthesis planning software that can propose fully-specified synthetic routes to target molecules.

In this talk, we will describe our recent efforts to develop such software. The overall synthesis planning workflow contains a number of interconnected modules. We focus on two critical aspects of computer-aided organic synthesis and how machine learning and other data-driven techniques have enabled new approaches to both challenges. First, we discuss the problem of retrosynthetic planning and how the recursive expansion and search strategy are both conducive to machine learning approaches. Second, we discuss the challenge of in silico reaction validation, which can be addressed by solving the inverse problem of forward reaction prediction. We summarize neural network-based approaches we have taken to develop models that can anticipate the products of a chemical reaction after being trained on previously published reactions. Finally, we describe how these techniques for retrosynthesis and forward prediction are integrated into an overall workflow that, for a given molecular target, predicts a rank ordered list of reaction paths that connect the target to purchasable starting materials via a series of plausible reaction steps.

(No Image Selected)